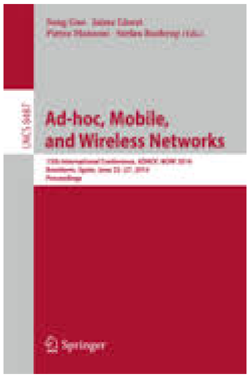


Article title: A Linear Programming Approach for K-Resilient and Reliability-Aware Design of Large-Scale Industrial Networks

Béla Genge, Piroska Haller, István Kiss
Petru Maior University of Tg. Mureș, Department of Informatics, Tg. Mureș, Romania
Email: bela.genge@ing.upm.ro, phaller@upm.ro, istvan.kiss@stud.upm.ro

This paper should be cited as:

B. Genge, P. Haller, I. Kiss: *A Linear Programming Approach for K-Resilient and Reliability-Aware Design of Large-Scale Industrial Networks*, International Conference on Ad Hoc, Mobile, and Wireless Networks (ADHOC-NOW), Lecture Notes in Computer Science, Volume 9143, Athens, Greece, pp. 288-302, 2015, DOI: 10.1007/978-3-319-19662-6_20.

The final publication is available at: http://link.springer.com/chapter/10.1007%2F978-3-319-19662-6_20

A Linear Programming Approach for K-Resilient and Reliability-Aware Design of Large-Scale Industrial Networks

Béla Genge, Piroska Haller, and István Kiss

"Petru Maior" University of Tîrgu Mureş
N. Iorga, No. 1, Tîrgu Mureş, Mureş, Romania, 540088
bela.genge@ing.upm.ro,phaller@upm.ro,istvan.kiss@stud.upm.ro

Abstract. The profound transformation of large-scale Industrial Control Systems (ICS), e.g., smart energy networks (Smart Grids), from a proprietary and isolated environment to a modern architecture brings several new challenges. Nowadays, ICS network designers need to accommodate a variety of devices and communication media/protocols with industry-specific requirements pertaining to real-time delivery of data packets, reliability, and resilience of communication networks. Therefore, this work proposes a novel network design methodology formulated as a Mixed Integer Linear Programming (MILP) problem. The developed problem accounts for different data flows routed across an overlay network of concentrators and embodies traditional ICS design requirements defined as linear constraints. Furthermore, the MILP problem defines a K-resilience factor to ensure the installation of K back-up paths, and a linear reliability constraint adapted from the field of fuzzy logic optimization. Experimental results demonstrate the efficiency and scalability of the proposed MILP problem.

Keywords: Industrial Control Systems, Energy Networks, Smart Grid, Linear Programming, Network Design, Resilience, Reliability

1 Introduction

Industrial Control Systems (ICS) have undergone a profound transformation from a traditional, isolated environment to a modern industrial architecture. Nowadays, ICS encompass a wide variety of devices, software, and communication protocols which paved the way towards the implementation of new features and operational paradigms such as smart electricity networks, more generally known as Smart Grid.

This ongoing transformation is a consequence of recognizing the importance of these infrastructures which are essential to the smooth functioning of our society. Due to their significant role, large-scale infrastructures such as smart electricity grids, transportation systems, oil and gas pipelines, are often characterized as Critical Infrastructures (CI). Without any doubt, one of the most debated and globally recognized CI is the electricity grid. In this direction the importance of upgrading the current infrastructure and its transformation towards the next electricity grid has been publicly acknowledged by academia, policy makers, and private companies as well. In this respect the US Department of Energy (DoE) [1] expressed the need to accelerate the deployment of advanced communication, control, and generally speaking the integration of modern Information and Communication Technologies (ICT), aiming at the modernization of the electricity grid. Subsequently, the DoE has set several performance targets for 2030, among which the reduction of peak energy demands, the efficient utilization of assets, and the integration of renewable energy resources. In Europe, on the other hand, we find several programs aiming at upgrading the current infrastructure and integrating novel Smart Grid applications [2, 3].

In light of these advancements, however, ICS network designers are faced with new challenges which need to accommodate a variety of devices and communication protocols on one hand, and industry-specific requirements on the other hand. In this respect, in the list of common ICS-specific design requirement parameters we find real-time delivery of data packets, reliability of communications, and communications resilience. These parameters are well-defined by international regulators/standardization bodies in the field of ICS. The National Institute of Standards and Technology (NIST) working group on Smart Grid [4], for instance, has identified the requirements for each of the aforementioned ICS parameters in the context of specific Smart Grid applications. Examples in this sense include the reliability of communications between 98% and 99.5%, and a latency of less than 10s for electric vehicle charging applications.

Based on these issues this work proposes a comprehensive methodology to design communication networks for large-scale ICS. The network design is formulated as a Mixed Integer Linear Programming (MILP) problem that accounts for data flows between different cyber assets, as well as traditional ICS design requirements pertaining to real-time traffic, reliability, and resilience. The developed methodology aims at harmonizing reliability and resilience aspects of ICS network design with well-known quality of service prerequisites, e.g., network latency, in order to deliver a modern communication infrastructure. By doing so, ICS network designers may integrate various design requirements into a methodology that minimizes the installation costs while ensuring that critical ICS-specific requirements are satisfied. It is also noteworthy that the proposed methodology may be used together with various other techniques as well. In this respect, for instance, the outcome of risk assessment given as a set of critical assets, may provide the necessary input to the methodology proposed in this paper. This way, a special emphasis may be placed on critical cyber assets in order to ensure an increased resilience and reliability factor to communications targeting such assets.

In particular, the proposed technique encompasses a *K-resilience* configuration factor to enable the deployment of K back-up communication paths for each main communication path. In this respect, for each data flow we define a *resilient communication group* (RCG), which encompasses the main data flow's communication path as well as all of its associated back-up communication paths. Reliability constraints are adopted from the traditional field of system safety, and are adapted to the present problem through the linear transformations proposed by Chiang *et al.* in the field of *fuzzy logic* optimization [5]. The proposed MILP problem is experimentally evaluated from several perspectives. More specifically, we assess the effect of various costs, resilience, and reliability parameters, on the solutions generated by the MILP model.

The remainder of this paper is organized as follows. An overview of related work is presented in Section 2. The problem statement together with a detailed description of the proposed methodology are given in Section 3. Then, the technique is experimentally evaluated from several perspectives in Section 4. The paper concludes in Section 5.

2 Related Work

This section provides a brief overview of related studies in the field of resilience and reliability-aware network design for ICS.

In [6], Carro Calvo *et al.* developed a genetic algorithm for optimal ICS network partitioning. The approach followed the traditional principles of ICS network design aimed at maximizing intra-network communications, minimizing inter-network communications, and balancing the communication over the resultant sub-networks. In [7], Zhou *et al.* formulated an optimization problem for ICS network design, which assumed a hierarchical switch-based ICS topology and incorporated costs in terms of the number of switches and of port utilization rates, traffic load balancing, as well as real-time traffic requirements expressed in terms of delay. The work of Zhang *et al.* in [8] focused on optimal ICS network design from the perspective of minimizing network delays. A relative delay metric was adopted to minimize communication delays with respect to the maximum tolerable delay. The same authors formulated in [9] an optimization problem which additionally incorporated ICS network reliability requirements in the form of probability of link failure. In [10], controllability analysis was used to reduce the size of the network design problem in water distribution systems by identifying the sub-set of nodes that fully cover the monitored physical process. Finally, the work of Zahidi *et al.* [11] showed that modern Integer Linear Programming (ILP) solvers are able to handle large network topologies. The effectiveness of ILP was demonstrated in optimizing the formation of clusters in Mobile Ad-Hoc Networks (MANETs).

Despite the variety of existing approaches the design problem of large-scale ICS networks has not been properly addressed yet. In the domain of the aforementioned studies, the approach presented in this paper is mostly related to the work of Zhang *et al.* [8, 9]. However, our methodology is more suited to large-scale ICS design due to the special emphasis on traffic flows and concentrator site installations. Additionally, we propose a K-resilience factor, which is not available in related work, and we adapt the non-linear reliability constraints to

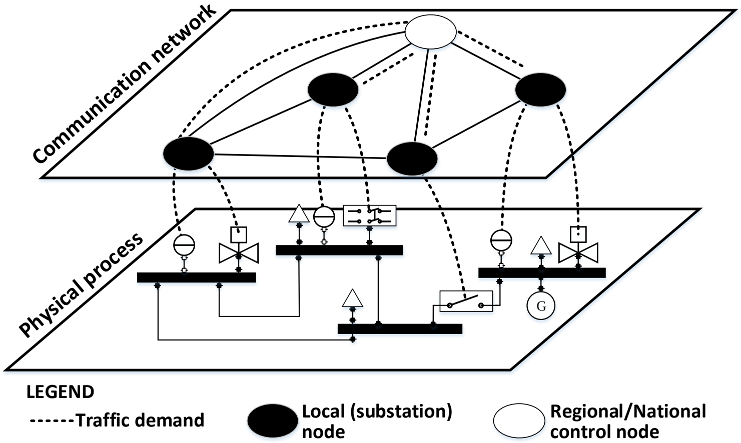


Fig. 1: Cyber-physical architecture of large-scale Industrial Control Systems.

the present linear problem by applying a transformation originating from the field of *fuzzy logic* optimization [5].

3 Proposed Network Design Model

This section presents the proposed network design problem. It starts with the problem statement and it continues with the definition of basic sets and symbols. Then, it provides a description of the objective function and of MILP constraints.

3.1 Problem Statement

We assume the cyber-physical architecture of a large-scale ICS as illustrated in Fig. 1. The central focus of this architecture are data exchanges between different end-points, depicted with dotted lines. Data flows, hereinafter called *traffic demands*, have a well-defined purpose. Their role is to enable the implementation of various control loops ranging from local actuation strategies, to large-scale / global decision-making algorithms. In the specific context of Smart Grid, for example, traffic demands ensure the timely delivery of price information announcements in smart customer energy management applications, the transfer of smart meter data to the utility, as well as the implementation of complex demand response programs aiming at a more effective energy planning and ultimately to the stability of the energy grid.

In this architecture traffic demands are routed across a hybrid hierarchical-meshed topology consisting of data concentrator nodes. These are well-known elements in the architecture of smart energy networks, and are mainly destined to aggregate and forward data from/into various communication mediums. Data concentrators may be implemented in various locations to aggregate data from various data networks. In this respect, at the home level, data concentrators aggregate traffic from smart meters, and different other devices specific to Smart Grid applications. Then, for each building data is aggregated from several different home networks, while for each neighborhood, data is aggregated from several building area networks. Subsequently, data is then aggregated from several networks at substation level in order to ensure that data is available to the utility and to customers.

Within this complex cyber-physical architecture ICS network design requirements need to account for various parameters such as: real-time delivery of data packets, reliability of communications, and communications resilience. Real-time requirements in ICS are defined for the majority of data flows associated with critical control loops. For instance, a common requirement is to ensure a strict packet delivery time for high-speed messages, e.g., alerts, in a certain range, e.g., in the 2ms to 10ms interval [12]. These limitations are necessary to ensure normal functioning of critical equipment and implementations need to ensure that these time limitations are met. On the other hand, reliability of communications is defined as the probability of successful communication over a specific time period [13]. Reliability is usually achieved by implementing redundant communication links and nodes, but also by increasing the *mean time to failure* of hardware. Finally, resilience is the ability of the system to function and to alleviate the disruptive impact of disturbances in various stress scenarios. For example, disruptive cyber attacks may cause significant packet delays and might block the delivery of critical commands to control hardware. Therefore, the resilience property might be implemented with alternative communication paths. Note that the aforementioned properties are not necessarily independent since for each resilience-assurance communication path network designers need to meet reliability and real-time requirements in order to comply with standard industry-specific prerequisites [12]. On the other hand, the difference between reliability and resilience is subtle, yet of paramount importance. In this respect, while reliability applies to normal up-time operation, resilience is formulated for conditions of stress, i.e., disruptive cyber attacks. Therefore, we believe that ICS network design methodologies need to incorporate not only reliability, but also resilience in order to build survivable ICS installations.

To accommodate these requirements, this work proposes a mathematical model of the ICS network design as a MILP problem. In the proposed MILP problem traffic demand paths, end-points, and concentrator sites are installed in such a way to minimize the costs of the infrastructure, while ensuring that constraints regarding real-time packet delivery, reliability, and K-resilience are satisfied.

3.2 Preliminary Notations

We define $I = \{1, 2, ..., i, ...\}$ to be the set of traffic demands (TD), and $J = \{1, 2, ..., j, ...\}$ the set of potential concentrator sites (CS). Then, let c_j^S to be the cost associated with installing CS j, c_{jl}^B be the cost of buying one unit of bandwidth between CSs j and l, c_{ij}^{A} the cost of buying one unit of bandwidth between the access end-point of TD i and CS j, and c_{ji}^{E} the cost of buying one unit of bandwidth between the egress end-point of TD i and CS j.

Then, let d_i denote the TD i, u_{jl} the link capacity between CSs j and l, and v_j the total capacity of combined access and egress demands for CS j. Considering the geographic location of access/egress end-points for each individual TD as well as the geographic location of CSs, the following binary connectivity parameters are defined. Let a_{ij} be a binary parameter with value 1 if the access end-point of TD i can connect to CS j, b_{jl} a binary parameter with value 1 if CS j can connect to CS l, and e_{ji} a binary parameter with value 1 if egress end-point of TD i can connect to CS j.

With respect to the ICS real-time performance requirements, the following communication latency parameters are defined. Let q_{ij}^{A} be the latency between access end-point of TD i and CS j, q_{ji}^{E} the latency between CS j and egress end-point of TD i, q_{jl}^{L} the latency between CSs j and l, and q_{j}^{C} the latency introduced by each CS j. Finally, we define the maximum tolerated latency for each individual TD i as q_{i}^{M}.

Next, we define the parameters concerning reliability. Let p_{jl} be the probability of failure of link j and l. p_{jl} is defined as a real number bounded between 0 and 1. The minimum desired path reliability for TD i is denoted by r_i.

The binary parameter h_{ik}, where $i, k \in I$, is defined such that $h_{ik} = 1$ if TDs i and k are part of the same RCG, and $h_{ik} = 0$, otherwise. Of particular importance for the proposed MILP problem is the configuration $h_{ii} = 1$, which reduces the complexity of constraints needed to identify TDs that are part of the same RCG. Before moving forward it is noteworthy that the proposed MILP problem assumes that primary and back-up TDs are part of the same set I. This is a significant design choice in the proposed MILP problem, which enables the applicability of the same constraints on both primary and back-up TDs. Moreover, it also facilitates the individual configuration of primary and back-up TDs which may represent a key aspect in the implementation of cost-efficient ICS communication networks.

Finally, we define g_j as a binary variable with value 1 if CS j is installed, x_{ij} as a binary variable with value 1 if the access end-point of TD i is connected to CS j, w_{ji} as a binary variable with value 1 if CS j is connected to the egress end-point of TD i, and t_{jl}^{i} as a binary variable with value 1 if TD i is routed on link (j, l), where $j, l \in J$.

3.3 Objective Function

The objective function aims at minimizing the cost of the installation:

$$\min \sum_{j \in J} c_j^S g_j + \sum_{j,l \in J, i \in I} c_{jl}^B t_{jl}^i d_i + \sum_{j \in J, i \in I} \left(c_{ij}^A x_{ij} + c_{ji}^E w_{ji} \right) d_i \tag{1}$$

The objective function (1) accounts for the total installation cost of concentrators, and communication links between CSs. In particular, the term $\sum c_j^S g_j$ is the total cost of installing all selected CSs, the term $\sum c_{jl}^{B} t_{jl}^{i} d_i$ is the total cost of bandwidth for routing traffic demands between CSs, and $\sum \left(c_{ij}^{A} x_{ij} + c_{ji}^{E} w_{ji} \right) d_i$ is the total cost of bandwidth for access and egress end-points connected to CSs.

3.4 Constraints

Next, we define the constraints pertaining to traffic demand paths, to real-time packet forwarding, to resilience, and finally, to reliability.

General demand flow and capacity constraints. The following constraints are defined to ensure the identification of traffic demand communication paths and to satisfy capacity restrictions.

$$\sum_{j \in J} x_{ij} = 1, \sum_{j \in J} w_{ji} = 1, \qquad \forall i \in I \qquad (2)$$

$$x_{ij} \leq a_{ij} g_j, w_{ji} \leq e_{ji} g_j, \qquad \forall i \in I, j \in J \qquad (3)$$

$$\sum_{i \in I} d_i \left(x_{ij} + w_{ji} \right) \leq v_j, \qquad \forall j \in J \qquad (4)$$

$$\sum_{i \in I} d_i t_{jl}^{i} \leq u_{jl} b_{jl} g_j, \sum_{i \in I} d_i t_{jl}^{i} \leq u_{jl} b_{jl} g_l, \qquad \forall j, l \in J \qquad (5)$$

$$x_{ij} - w_{ji} - \sum_{l \in J} \left(t_{jl}^{i} - t_{lj}^{i} \right) = 0, \qquad \forall j \in J, i \in I \qquad (6)$$

The constraints defined in Eq. (2) and (3) limit the number of connections between access/egress demand end-points and CSs to exactly one. Constraints defined in Eq. (4) and (5) impose restrictions with respect to concentrators and link capacities. Finally, constraints defined in Eq. (6) are classical multicommodity flow conservation equations [14].

Real-time communication constraints. The constraints formulated as Eq. (7) force the selection of routing paths that fulfill the latency requirements defined for each demand. In particular, for each demand i, the term $\sum \left(q_{ij}^{A} x_{ij} + q_{ji}^{E} w_{ji} \right)$ is the sum of latencies for access and egress links, the term $\sum t_{jl}^{i} \left(q_{jl}^{L} + q_{j}^{C} \right)$ is the sum of latencies owed to CSs and links between CSs, and the term $\sum q_{j}^{C} w_{ji}$ is the latency of the egress CS.

$$\sum_{j \in J} \left(q_{ij}^{A} x_{ij} + q_{ji}^{E} w_{ji} \right) + \sum_{j,l \in J} t_{jl}^{i} \left(q_{jl}^{L} + q_{j}^{C} \right) + \sum_{j \in J} q_{j}^{C} w_{ji} \leq q_{i}^{M}, \qquad \forall i \in I \qquad (7)$$

K-resilience constraints. As previously mentioned, in this work the *K-resilience* property of ICS communications is achieved by expanding set I with K back-up traffic demands for each primary TD. The association of TDs in a specific

RCG is achieved by means of parameter h_{ik}. In this respect we define two distinct resilience cases. In the first case we assume the provisioning of independent communication links in a specific RCG. This ensures resilience to communication link failure, however, it does not provide resilience in case of compromised/failed concentrator nodes. Therefore, a second case is defined in which both links and concentrator nodes are distinctly chosen in each RCG.

The MILP problem at hand supports link and node independency by means of additional linear constraints. In this respect, Eq. (8) defines constraints pertaining to access and egress TD link independency, while Eq. (9) defines link independency between CSs selected in the same RCG.

$$\sum_{k \in I} x_{kj} h_{ik} \leq 1, \sum_{k \in I} w_{kj} h_{ik} \leq 1, \qquad \forall i \in I, j \in J \tag{8}$$

$$\sum_{k \in I} (t^k_{jl} + t^k_{lj}) h_{ik} \leq 1, \qquad \forall i \in I, j, l \in J \tag{9}$$

Conversely, the proposed MILP problem imposes concentrator node and link independency simultaneously through the linear constraints defined in Eq. (10). These include stricter conditions than the ones defined in Eq. (9). Consequently, if node independency is needed, the proposed MILP problem shall include the constraints defined in Eq. (10), instead of the ones deinfed in Eq. (9).

$$\sum_{k \in I, l \in J} (t^k_{jl} + t^k_{lj}) h_{ik} \leq 1, \qquad \forall i \in I, j \in J \tag{10}$$

Reliability constraints. Reliability in constrain-oriented programming is a well-known problem [15, 16]. For the MILP at hand the reliability of communication path for a specific TD i is defined as:

$$\prod_{j,l \in J} (1 - p_{jl} t^i_{jl}) \geq r_i, \qquad \forall i \in I \tag{11}$$

Unfortunately, this constraint is non-linear due to the multiplication of t^i_{jl} variables. However, based on the observation that t^i_{jl} is a binary variable, the transformation defined in [5] can be applied to derive a linear constraint. More specifically, owed to the binary range of values for t^i_{jl}, it follows that $(1 - p_{jl} t^i_{jl}) = (1 - p_{jl})^{t^i_{jl}}$. Consequently, constraint (11) can be rewritten as:

$$\prod_{j,l \in J} (1 - p_{jl})^{t^i_{jl}} \geq r_i, \qquad \forall i \in I \tag{12}$$

Then, by applying the natural logarithm function on both sides of inequality (12) we obtain the following linear reliability constraint, which is adopted in the proposed MILP problem:

$$\sum_{j,l \in J} t^i_{jl} \mathtt{ln}(1 - p_{jl}) \geq \mathtt{ln}(r_i), \qquad \forall i \in I \tag{13}$$

4 Experimental Results

In this section we provide experimental results on the applicability of the proposed MILP problem to the design of large-scale ICS communication networks.

4.1 Experimental Scenario and Parameters

The experimental assessment is conducted in the context of electricity grid communication networks. As a reference model we assume the IEEE 30-bus model [17] consisting of 6 generators and 30 substations.

The communication infrastructure of large-scale electricity grid networks is composed of several independently governed communication networks. In general, utilities rely on a main, internal communication network in order to supervise and control large-scale physical processes. However, regulations impose the installation of at least one back-up communication line leased from a different Internet Service Provider (ISP) with specific Quality of Service requirements [18]. In this respect, the problem at hand assumes that the communication medium between concentrator nodes (including the concentrator nodes) may be provided by two different ISPs: ISP_1 (primary communication network), and ISP_2 (secondary communication network). We further assume the following parameter values. We assume a first, primary communication network (governed by ISP_1) where CSs are installed in the premises of substations, which yield a number of 30 CSs. Next, we assume the same number of CSs in the second communication network (governed by ISP_2). We further assume 30 main traffic demands originating from each substation, which need to be routed to regional/national monitoring nodes connected exclusively to the second communication network.

In order to ensure a realistic estimation of parameters we arranged meetings with representatives of a local electricity grid operator and a local ISP provider. Discussions confirmed the validity of the aforementioned assumptions on communication infrastructure, and guided the choice of parameter values pertaining to costs and traffic magnitude. For the provisioning of CSs we initially assume that $c_j^S = 30$ monetary units (MU) for all CSs, while the cost of one bandwidth unit is assumed to be $c_{jl}^B = 3$MU per Mb/s, and $c_{ij}^A = c_{ji}^E = 1$MU/Mb/s. We further assume that the traffic that needs to be routed for each TD is $d_i = 100$Mb/s, and the capacity of each CS is $v_j = 8000$Mb/s and $u_{jl} = 1000$Mb/s.

The initial connectivity between CSs is randomly configured through a uniform distribution function such that $b_{jl} = 1$ with a probability of 20% if both CSs j and l are in the primary network and with a probability of 5% if at least one CS is in the secondary network. The initial connectivity of access TD end-point to CSs is similarly configured such that $a_{ij} = 1$ with a probability of 20% if CS j is in the primary network, and with a probability of 5%, otherwise. Since the egress end-point of TDs are exclusively assumed to be connected to CSs in the secondary network, we assume that $e_{ji} = 1$ with a probability of 20%.

With respect to latency, for the sake of simplicity we assume that $q_{ij}^A = q_{ji}^E = q_{jl}^L = q_j^C = 1$ms. Subsequently, we assume that the maximum tolerated latency

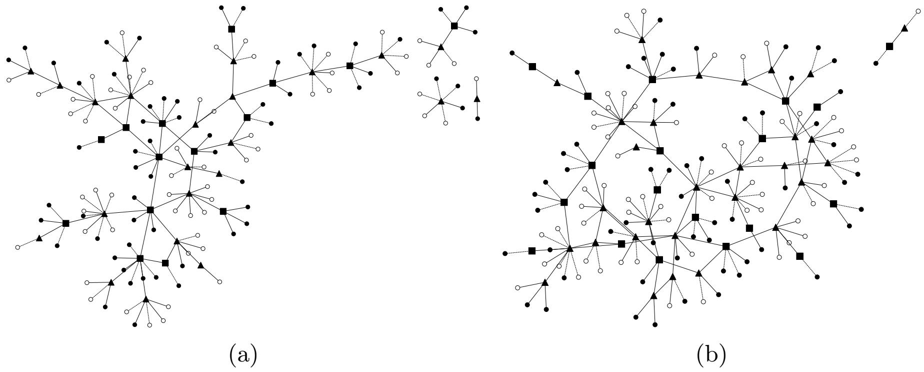


Fig. 2: Effect of costs on the generated solution: (a) equal costs in the primary and secondary networks; and (b) five times higher costs in the primary network. CSs in the primary and in the secondary network are represented with boxes and triangles, respectively. Access and egress end-points are represented by filled and empty circles, respectively.

is of $q_i^M = 10$ms [12]. The probability of failure is assumed to be of 0.01%, i.e., $p_{jl} = 0.0001$, while the minimum required reliability is assumed to be of 99.8%, i.e., $r_i = 0.998$ [18]. Finally, the resilience communication group parameters h_{ik} are initialized according to the values of K, as discussed later in the following sections. Initially, we assume that $K = 1$.

The network design problem was implemented in AIMMS [19], where we adopted the popular CPLEX engine as a MILP solver. All experiments were performed on a Windows 7 OS host, with Pentium Dual Core CPU at 3GHz, and 4 GB of memory.

4.2 Effect of Costs on the Generated Solution

Obviously, the focus of the objective function on costs will yield an increased sensitivity of the generated solutions on the components' costs. An example in this sense is provided in Fig. 2a and 2b. Here it can be seen that the increase of components' costs in the primary network, in comparison to the costs of components in the secondary network, triggers profound changes in the architecture of the generated solutions. As a result, more CSs are selected from the secondary network, which ensures minimum provisioning costs for the generated solution.

This particular trend is more visible from numerical results. In this respect, we performed several tests in which we increased the costs of components in the primary network starting from a ratio of 0.5 with respect to the cost of components in the secondary network, and up to a ratio of 5. Given the probabilistic configuration of connection options, for each cost value we ran the AIMMS solver 10 times. As illustrated in Fig. 3, the increase in the price ratio between the primary and the secondary network decreases the number of CSs allocated from the

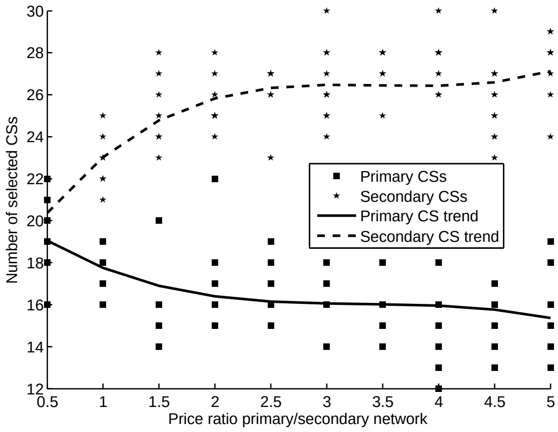
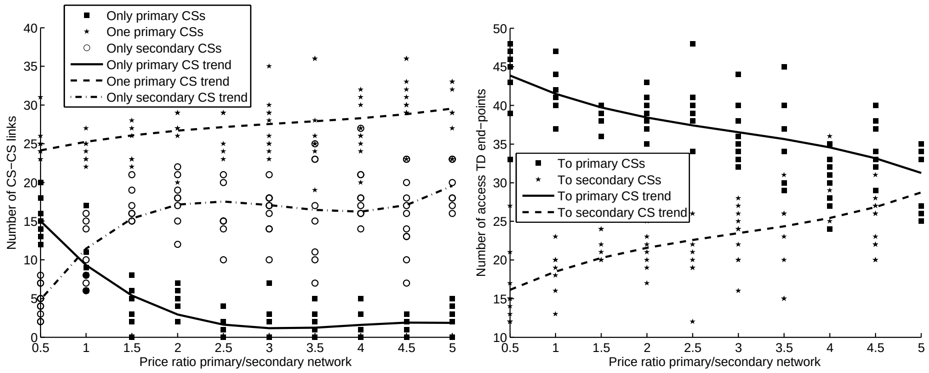


Fig. 3: Effect of increasing the costs of infrastructure in the primary network on the selection of CSs.



(a)

(b)

Fig. 4: Effect of increasing the costs of infrastructure in the primary network on the selection of links: (a) between CSs located in the primary and/or in the secondary network; and (b) between access TD end-points and CSs located in the primary or in the secondary network.

primary network, and it increases at the same time the number of CSs allocated from the secondary network.

A similar behavior was measured for the allocation of links between different network components, i.e., CSs and TDs (see Fig. 4a and 4b). Here (Fig. 4a) it can be seen that the increase in the price ratio will increase the preference of linking together more CSs from the secondary network and less from the primary network. The same trend is also visible for the connection of access TD end-points to CSs (see Fig. 4b).

Table 1: Percentage of feasible configurations as a function of K and α

	α [%]									
K	5	10	15	20	25	30	35	40	45	50
1	0%	0%	0%	**81%**	**91%**	**100%**	**100%**	**100%**	**100%**	**100%**
2	0%	0%	0%	**23%**	**70%**	**98%**	**100%**	**100%**	**100%**	**100%**
3	0%	0%	0%	0%	**33%**	**81%**	**90%**	**100%**	**100%**	**100%**
4	0%	0%	0%	0%	0%	**41%**	**80%**	**100%**	**100%**	**100%**
5	0%	0%	0%	0%	0%	**15%**	**52%**	**91%**	**97%**	**100%**

4.3 Effect of K-resilience Factor on the Generated Solution

Until this point we assumed the implementation of only one back-up path for each TD, i.e., $K = 1$. However, the requirement of link and node independence between different TDs of the same RCG, may yield an unfeasible linear problem for which the solver cannot provide a solution. This scenario holds especially when the configuration does not provide sufficient connectivity options in parameters a_{ij}, b_{jl}, or e_{ji}. Since the tests presented in this work assume a probabilistic configuration of connectivity parameters, this sub-section evaluates the impact of the number of back-up paths (K) and of different probability distribution values (denoted by α) on the problem's feasibility. For this purpose we assume that $\alpha \in \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$ (in percentage), and $K \in \{1, 2, 3, 4, 5\}$. For each (α, K) configuration we run the AIMMS solver 40 times.

The measured percentage of feasible solutions for each (K, α) configuration is given in Table 1. Here it is shown that for $\alpha \leq 15\%$ the solver is unable to find a feasible solution for any values of K. This is explained by the link independence constraint which requires that access, egress, and CS-CS links to be independent within a specific RCG. By increasing the value of α above 15%, however, for $K = 1$, the solver finds that more than 81% of configurations lead to a feasible solution. Nevertheless, by increasing the value of K, that is, by expanding the problem with K additional TDs in each RCG, the percentage of feasible solutions decreases dramatically. As an example in this sense, for $\alpha = 30\%$ and $K = 1$ we measure a 100% rate of feasible solutions, while for the same α and $K = 5$ the rate of feasible solutions drops to 15%. These results are a clear indication on the need to carefully formulate the MILP problem, since minor changes to input data may lead to significantly different outputs.

4.4 Effect of Reliability Parameters on the Generated Solution

The reliability constraints impose the identification of solutions that satisfy the minimum path reliability condition. In this section we assume a minimum reliability value of 99.8%, that is $r_i = 0.998$ for all $i \in I$. As a result, the generated solutions guarantee that for all TDs the communication paths have a reliability of at lest 99.8%. In Fig. 5 we illustrate the average reliability of communication paths for several solver runs. Here the value of K was fixed to 1, the value of α

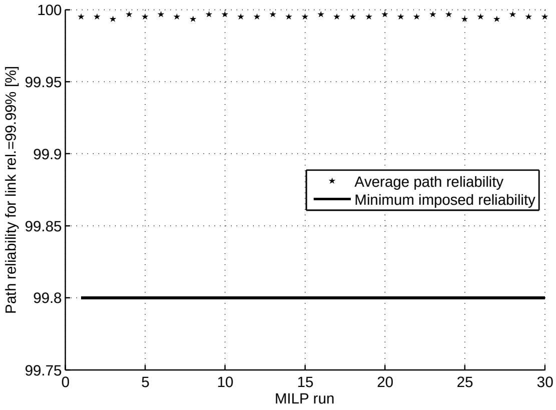


Fig. 5: The average reliability of communication paths considering that the reliability of each individual j, l link is 99.99%.

was fixed to 20%, while individual link reliability was of 99.99%. As a result, the measured path reliability for each solution was well above the minimum of 99.8%. In fact, we measured a minimum path reliability of 99.993%, and a maximum path reliability of 99.996%. Obviously, the path reliability is directly influenced by the individual link reliability values. However, it is noteworthy that if the configuration is feasible, then the proposed MILP problem will always yield a cost-optimal solution that guarantees that reliability constraints are satisfied.

4.5 Execution Time

We performed several experiments to test the execution time of the CPLEX solver on the proposed MILP problem. As tabulated in Table 2, the execution time is influenced by the problems' dimension in terms of the number of CSs, the number of TDs, and the number of configured connection options, i.e., the magnitude of α. In this respect for $\alpha = 20\%$ and for 10 CSs the solver determined that there are no feasible configurations available. Nevertheless, by increasing α to 40% solutions are generated in 0.2s for 10 CSs and 10 TDs, and in 0.62s for 10 CSs and 60 CSs. By further increasing the number of CSs we also notice a linear increase in the solvers' execution time. Nevertheless, as depicted in Table 2 the execution time for 50 CSs and 60 TDs does not exceed 30s, while for 100 CSs and 60 TDs the execution time is below 160s, i.e., below 3 minutes. These results are a clear indication on the scalability and applicability of the proposed MILP problem to large-scale ICS networks. The linear increase of the execution time provides further confirmation on the applicability of constraint-based programming to the design of large-scale ICS network.

Table 2: Execution time (s) for $\alpha = 20\%$ and $\alpha = 40\%$ (we assume $K = 1$). Unfeasible configurations are denoted by '–'.

	Total TD count including back-up					
CS count	10	20	30	40	50	60
10	–/0.2	–/0.29	–/0.32	–/0.41	–/0.53	–/0.62
20	0.36/0.37	0.64/0.73	0.99/1.19	–/1.5	–/2.05	–/2.3
30	0.68/0.88	1.45/1.67	2.11/2.61	3.01/3.58	3.61/5.72	4.37/5.93
40	1.28/1.53	2.54/3.28	3.88/5.38	5.29/8.20	6.81/11.4	8.14/12.2
50	2.02/2.49	4.04/5.66	6.17/8.80	8.77/15.8	10.5/21.1	13.3/27.7
60	3.09/3.68	6.48/8.52	9.72/14.1	13.0/24.7	17.2/35.7	21.1/43.2
70	4.04/5.56	8.28/12.2	13.4/23.0	17.7/33.6	23.3/47.6	28.2/55.3
80	5.73/7.50	11.0/15.9	17.1/32.7	25.4/45.5	30.7/58.9	38.5/79.3
90	7.12/10.1	14.4/22.2	22.3/42.2	30.2/59.3	38.8/79.3	48.0/100.7
100	9.03/12.8	18.1/30.7	28.5/54.5	39.3/77.1	50.5/101.2	63.8/158.8

5 Conclusion

We proposed a novel MILP problem that accounts for the provisioning of traditional concentrator equipment, of traffic flows routed across an overlay network, and most importantly of typical ICS design requirements pertaining to real-time packet delivery, communications reliability, and resilience. The approach saved costs on investments in the system's resources, enhancing at the same time the resilience of ICS networks by a factor of K through the deployment of back-up communication paths. Experimental results proved that the technique is scalable and applicable to large-scale ICS such as modern energy network. As future work we intend to further expand the developed MILP with security requirements in order to deliver a methodology that ensures security and resilience-aware design of ICS communication networks.

Acknowledgments. This research was supported by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme.

References

1. U.S. Department of Energy: Smart grid research and development: Multi-year program plan (mypp) 2010-2014 (September 2012)
2. European Commission: European smart grids technology platform, eur 22040 (September 2006)
3. European Commission: Horizon 2020 energy calls (January 2015)
4. NIST: NIST Smart Grid Collaboration. `http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/WebHome` (2011)
5. Chiang, C., Hwang, M., Liu, Y.: An alternative formulation for certain fuzzy set-covering problems. Mathematical and Computer Modelling **42**(34) (2005) 363 – 365

6. Carro-Calvo, L., Salcedo-Sanz, S., Portilla-Figueras, J.A., Ortiz-Garca, E.: A genetic algorithm with switch-device encoding for optimal partition of switched industrial ethernet networks. Journal of Network and Computer Applications **33**(4) (2010) 375 – 382
7. Zhou, Z., Chen, B., Wang, H., Fan, Z.: Study on the evolutionary optimisation of the topology of network control systems. Enterprise Information Systems **4**(3) (2010) 247–264
8. Zhang, L., Lampe, M., Wang, Z.: A hybrid genetic algorithm to optimize device allocation in industrial ethernet networks with real-time constraints. Journal of Zhejiang University SCIENCE C **12**(12) (2011) 965–975
9. Zhang, L., Lampe, M., Wang, Z.: Multi-objective topology design of industrial ethernet networks. Frequenz **66**(5-6) (2012) 159–165
10. Diao, K., Rauch, W.: Controllability analysis as a pre-selection method for sensor placement in water distribution systems. Water Research **47**(16) (2013) 6097 – 6108
11. Zahidi, S., Aloul, F., Sagahyroon, A., El-Hajj, W.: Optimizing complex cluster formation in manets using sat/ilp techniques. Sensors Journal, IEEE **13**(6) (June 2013) 2400–2412
12. Institute of Electrical and Electronics Engineers: IEEE 1646-2004 standard: communication delivery time performance requirements for electric power substation automation (2004)
13. IEC62439: Industrial communication networks - high availability automation networks. (2012)
14. Meixner, C., Dikbiyik, F., Tornatore, M., Chuah, C., Mukherjee, B.: Disaster-resilient virtual-network mapping and adaptation in optical networks. In: Optical Network Design and Modeling (ONDM), 2013 17th International Conference on. (April 2013) 107–112
15. Hsieh, Y.C.: A linear approximation for redundant reliability problems with multiple component choices. Computers & Industrial Engineering **44**(1) (2003) 91 – 103
16. Billionnet, A.: Redundancy allocation for series-parallel systems using integer linear programming. Reliability, IEEE Transactions on **57**(3) (Sept 2008) 507–516
17. University of Washington - Electrical Engineering: Power Systems Test Case Archive. `http://www.ee.washington.edu/research/pstca/` (2015)
18. Nazionale, G.R.T.: Criteri di connessione al sistema di regolazione della tensione del GRTN. Terna, Italy (2005)
19. AIMMS: Advanced Interactive Multidimensional Modeling System. `http://www.aimms.com/aimms/` (2014)